

Supplementary I: type I error rate and power comparison

Type I error rate comparison:

To compare false positive (type I) error rates of the EGHST, HMMST and EHHST, we simulated data under the null of the EGHST statistic. We simulated genotype data using the SNP allele frequencies from the CHB+JPT and YRI samples on chromosomes 1, 2, and 15, assuming HWE and linkage equilibrium.

Table 1 summarized the performance of the EGHST statistic over 10^8 random samples of $n = 60, 90, 125, 250, 500,$ and 1250 individuals. The results in Table 1 suggested that false positive rates were appropriate for $\alpha = 0.05$ and 0.01 when $n \geq 60$. For $\alpha = 0.001$, false positive rates were too high when $n \leq 250$ and were close to the nominal level when $n \geq 500$. Finally for $\alpha = 0.0001$, it took a sample size of at least $n = 1250$ to get an accurate false positive rate.

Table 2 reported false positive rates for the HMMST with data simulated under the same conditions. We also added results for the CEU sample. Under the unrealistic null hypothesis appropriate to the EGHST, the HMMST was too conservative. False positive rates were smaller than the nominal level for $\alpha = 0.05, 0.01, 0.001, 0.0001,$ and 0.00001 when sample size was 60 or greater. For $\alpha = 0.000001$, false positive rates were reasonable when $n \geq 125$.

Table 3 reported false positive rates for the EHHST. Since the high computational demanding of EHHST, 10^5 simulations instead of 10^8 were performed to calculate a type I error rate. False positive rates were smaller than the nominal level for $\alpha = 0.05, 0.01,$ and 0.001 when sample size was 60 or greater. At the nominal level $\alpha = 0.0001$, the false positive rates were reasonable when sample size was 60 or greater. Comparing with results of Table 1, the false positive rates for the EHHST were lower than those of the EGHST. At the nominal level of 0.05 , the false positive rates for the EHHST were lower than those of the HMMST.

In conclusion, the EHHST was the most conservative one among the three tests.

Power comparison:

To make power comparison for the three tests proposed, we simulated data under the same conditions as those of Table 3 in the main text by SelSim. Again, 5,000 samples of 200 chromosomes (i.e., $n = 100$ individuals) or 120 chromosomes (i.e., $n = 60$ individuals) were simulated to calculate the empirical powers for EGHST and HMMST. The results were reported in Table 4 and Table 5. As expected, the power of EGHST is higher than that of HMMST which is generally more powerful than EHHST.

Table 1: Type I error rates of the extended genotype-based homozygosity score test (EGHST). All results based on 10^8 simulations and the HapMap Phase II SNP allele frequencies.

Sample Size n	Chromosome	Population - Allele Frequency Based	Type I Error Rates When			
			$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0.0001$
60	chromosome 1	CHB+JPT	0.04938718	0.01035739	0.00135778	0.00022274
		YRI	0.04937440	0.01028255	0.00132343	0.00021185
	chromosome 2	CHB+JPT	0.04940133	0.01034034	0.00135832	0.00022680
		YRI	0.04947162	0.01030993	0.00132751	0.00021625
	chromosome 15	CHB+JPT	0.04932188	0.01031996	0.00135838	0.00022544
		YRI	0.04940568	0.01031936	0.00132825	0.00021288
90	chromosome 1	CHB+JPT	0.04952430	0.01020326	0.00123662	0.00018285
		YRI	0.04955716	0.01019437	0.00122318	0.00017691
	chromosome 2	CHB+JPT	0.04955975	0.01022223	0.00124360	0.00018541
		YRI	0.04964012	0.01020800	0.00123110	0.00017911
	chromosome 15	CHB+JPT	0.04960285	0.01022962	0.00124236	0.00018548
		YRI	0.04964285	0.01022434	0.00122597	0.00017911
125	chromosome 1	CHB+JPT	0.04968785	0.01014673	0.00117489	0.00016028
		YRI	0.04974835	0.01015954	0.00116450	0.00015561
	chromosome 2	CHB+JPT	0.04973875	0.01018076	0.00118329	0.00016391
		YRI	0.04964882	0.01011992	0.00116113	0.00015746
	chromosome 15	CHB+JPT	0.04966775	0.01012681	0.00116167	0.00015570
		YRI	0.04972462	0.01014865	0.00115829	0.00015481
250	chromosome 1	CHB+JPT	0.04983445	0.01006635	0.00108067	0.00012857
		YRI	0.04988641	0.01009465	0.00108552	0.00012828
	chromosome 2	CHB+JPT	0.04989234	0.01009633	0.00109923	0.00013297
		YRI	0.04989319	0.01007496	0.00108337	0.00012691
	chromosome 15	CHB+JPT	0.04979736	0.01008825	0.00109559	0.00013371
		YRI	0.04982615	0.01007379	0.00108525	0.00012952
500	chromosome 1	CHB+JPT	0.04988070	0.01003214	0.00104432	0.00011518
		YRI	0.04992955	0.01003731	0.00104643	0.00011390
	chromosome 2	CHB+JPT	0.04988924	0.01003926	0.00103772	0.00011406
		YRI	0.04992835	0.01004388	0.00103968	0.00011400
	chromosome 15	CHB+JPT	0.04995869	0.01001675	0.00103760	0.00011289
		YRI	0.04992122	0.01004642	0.00104140	0.00011416
1250	chromosome 1	CHB+JPT	0.04992160	0.01000125	0.00101399	0.00010860
		YRI	0.05001056	0.01002057	0.00101685	0.00010386
	chromosome 2	CHB+JPT	0.04990222	0.00998176	0.00102016	0.00010494
		YRI	0.04992216	0.01001738	0.00101428	0.00010371
	chromosome 15	CHB+JPT	0.04996471	0.01004044	0.00102895	0.00010903
		YRI	0.04991000	0.01000108	0.00101769	0.00010529

Table 2: Type I error rates of the hidden Markov model score test (HMMST). All results based on 10^8 simulations and the HapMap Phase II SNP allele frequencies.

Sample Size n	Chromosome	Population - Allele Frequency Based	Type I Error Rates When $\alpha =$					
			0.05	0.01	0.001	0.0001	0.00001	0.000001
60	Chromosome 1	CEU	0.0326	0.0040	0.00024	2.7e-05	5.5e-06	1.4e-06
		CHB+JPT	0.0337	0.0043	0.00029	3.9e-05	9.0e-06	2.8e-06
		YRI	0.0318	0.0038	0.00021	2.0e-05	3.0e-06	5.9e-07
	Chromosome 2	CEU	0.0324	0.0040	0.00025	2.9e-05	6.1e-06	1.9e-06
		CHB+JPT	0.0332	0.0043	0.00031	5.0e-05	1.5e-05	6.2e-06
		YRI	0.0317	0.0037	0.00020	2.0e-05	2.8e-06	4.9e-07
	Chromosome 15	CEU	0.0325	0.0040	0.00026	3.4e-05	7.9e-06	2.7e-06
		CHB+JPT	0.0331	0.0042	0.00028	4.0e-05	1.2e-05	6.3e-06
		YRI	0.0317	0.0037	0.00020	1.8e-05	2.6e-06	5.5e-07
90	Chromosome 1	CEU	0.0313	0.0041	0.00024	2.2e-05	2.9e-06	4.1e-07
		CHB+JPT	0.0327	0.0043	0.00027	2.7e-05	4.7e-06	1.3e-06
		YRI	0.0305	0.0039	0.00022	1.9e-05	2.4e-06	4.0e-07
	Chromosome 2	CEU	0.0315	0.0041	0.00025	2.3e-05	2.9e-06	4.5e-07
		CHB+JPT	0.0320	0.0043	0.00028	3.0e-05	5.3e-06	1.3e-06
		YRI	0.0305	0.0039	0.00022	1.8e-05	2.5e-06	4.1e-07
	Chromosome 15	CEU	0.0315	0.0042	0.00026	2.8e-05	4.8e-06	1.3e-06
		CHB+JPT	0.0320	0.0043	0.00027	2.9e-05	6.2e-06	2.4e-06
		YRI	0.0306	0.0039	0.00022	1.8e-05	2.3e-06	4.3e-07
125	Chromosome 1	CEU	0.0308	0.0042	0.00026	2.2e-05	2.6e-06	4.0e-07
		CHB+JPT	0.0315	0.0044	0.00028	2.4e-05	3.1e-06	5.6e-07
		YRI	0.0299	0.0039	0.00023	1.7e-05	1.9e-06	3.4e-07
	Chromosome 2	CEU	0.0310	0.0042	0.00026	2.2e-05	2.4e-06	3.5e-07
		CHB+JPT	0.0315	0.0044	0.00029	2.7e-05	3.3e-06	6.4e-07
		YRI	0.0299	0.0039	0.00023	1.8e-05	1.7e-06	2.1e-07
	Chromosome 15	CEU	0.0310	0.0043	0.00027	2.4e-05	2.7e-06	4.3e-07
		CHB+JPT	0.0315	0.0044	0.00028	2.7e-05	4.2e-06	1.2e-06
		YRI	0.0301	0.0040	0.00024	1.9e-05	2.2e-06	2.6e-07
250	Chromosome 1	CEU	0.0301	0.0043	0.00029	2.3e-05	2.2e-06	2.4e-07
		CHB+JPT	0.0307	0.0044	0.00030	2.5e-05	2.5e-06	3.4e-07
		YRI	0.0292	0.0041	0.00026	1.9e-05	1.6e-06	2.0e-07
	Chromosome 2	CEU	0.0305	0.0044	0.00030	2.3e-05	2.5e-06	2.6e-07
		CHB+JPT	0.0307	0.0045	0.00031	2.6e-05	2.7e-06	3.4e-07
		YRI	0.0293	0.0041	0.00026	2.1e-05	1.9e-06	2.5e-07
	Chromosome 15	CEU	0.0305	0.0044	0.00029	2.3e-05	2.2e-06	3.1e-07
		CHB+JPT	0.0309	0.0045	0.00031	2.5e-05	2.5e-06	2.5e-07
		YRI	0.0294	0.0041	0.00027	2.1e-05	2.3e-06	3.2e-07

Table 3: Type I error rates of the extended haplotype-based homozygosity score test (EHHST). All results based on 10^5 simulations and the HapMap Phase II SNP allele frequencies.

Sample Size n	Chromosome	Population - Allele Frequency Based	Type I Error Rates When $\alpha =$			
			0.05	0.01	0.001	0.0001
60	Chromosome 1	CEU	0.02271	0.00373	0.00047	0.00009
		CHB+JPT	0.02407	0.00441	0.00056	0.00012
		YRI	0.02331	0.00382	0.00053	0.00010
	Chromosome 2	CEU	0.02586	0.00447	0.00037	0.00010
		CHB+JPT	0.02561	0.00477	0.00078	0.00011
		YRI	0.02364	0.00344	0.00039	0.00006
	Chromosome 15	CEU	0.02349	0.00386	0.00041	0.00003
		CHB+JPT	0.02445	0.00404	0.00045	0.00008
		YRI	0.02308	0.00373	0.00034	0.00006
90	Chromosome 1	CEU	0.02326	0.00353	0.00032	0.00006
		CHB+JPT	0.02371	0.00389	0.00032	0.00001
		YRI	0.02209	0.00373	0.00046	0.00005
	Chromosome 2	CEU	0.02446	0.00387	0.00032	0.00006
		CHB+JPT	0.02433	0.00463	0.00060	0.00009
		YRI	0.02454	0.00366	0.00023	0.00003
	Chromosome 15	CEU	0.02336	0.00361	0.00049	0.00006
		CHB+JPT	0.02424	0.00367	0.00038	0.00008
		YRI	0.02324	0.00343	0.00030	0.00007
125	Chromosome 1	CEU	0.02323	0.00336	0.00023	0.00002
		CHB+JPT	0.02433	0.00394	0.00037	0.00005
		YRI	0.02208	0.00324	0.00027	0.00001
	Chromosome 2	CEU	0.02264	0.00355	0.00036	0.00006
		CHB+JPT	0.02379	0.00423	0.00054	0.00004
		YRI	0.02316	0.00296	0.00017	0.00004
	Chromosome 15	CEU	0.02362	0.00364	0.00032	0.00002
		CHB+JPT	0.02393	0.00416	0.00060	0.00008
		YRI	0.02345	0.00335	0.00028	0.00002
250	Chromosome 1	CEU	0.02259	0.00350	0.00017	0.00002
		CHB+JPT	0.02447	0.00392	0.00034	0.00004
		YRI	0.02149	0.00318	0.00017	0.00002
	Chromosome 2	CEU	0.02439	0.00334	0.00027	0.00002
		CHB+JPT	0.02380	0.00397	0.00026	0.00002
		YRI	0.02336	0.00348	0.00020	0.00001
	Chromosome 15	CEU	0.02325	0.00352	0.00022	0.00000
		CHB+JPT	0.02424	0.00379	0.00037	0.00006
		YRI	0.02200	0.00328	0.00029	0.00002

Table 4: Power of the Extended genotype-based homozygosity score test (EGHST). All results were based on 5,000 simulations using Software SelSim. The rows marked by # contain results which were calculated using the same models and parameters as those of Figure 1 of Hanchard et al. (2006). **Abbreviation:** Freq. — Frequency, Popu. — Population.

Selection Coefficient	Sample Size n	Recombination Rates ρ	Nominal Level α	Present Day Popu. Freq. of Derived Allele						
				0.1	0.2	0.4	0.6	0.8	0.9	
s=500	100	1.5#	0.05	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	
			0.01	0.9982	1.0000	1.0000	1.0000	1.0000	1.0000	
		3#	0.05	0.9962	1.0000	1.0000	1.0000	1.0000	1.0000	
			0.01	0.9796	1.0000	1.0000	1.0000	1.0000	0.9996	
		6#	0.05	0.9638	1.0000	1.0000	1.0000	1.0000	0.9970	0.9924
			0.01	0.8770	0.9996	1.0000	1.0000	1.0000	0.9970	0.9636
		9	0.05	0.9198	0.9996	1.0000	1.0000	1.0000	0.9998	0.9532
			0.01	0.7746	0.9986	1.0000	1.0000	1.0000	0.9994	0.8444
	60	1.5	0.05	0.9826	0.9990	1.0000	1.0000	1.0000	1.0000	1.0000
			0.01	0.9748	0.9976	1.0000	1.0000	1.0000	1.0000	1.0000
		3	0.05	0.9162	0.9904	1.0000	0.9996	1.0000	1.0000	0.9934
			0.01	0.8158	0.9820	1.0000	0.9996	0.9996	0.9996	0.9816
		6	0.05	0.7682	0.9930	1.0000	1.0000	1.0000	0.9972	0.9044
			0.01	0.5936	0.9724	1.0000	1.0000	1.0000	0.9856	0.7680
		9	0.05	0.6768	0.9786	1.0000	1.0000	1.0000	0.9784	0.7396
			0.01	0.4584	0.9336	0.9994	1.0000	1.0000	0.9264	0.5092

Table 5: Power of the hidden Markov model score test (HMMST). All results were based on 5,000 simulations using Software SelSim. The rows marked by # contain results which were calculated using the same models and parameters as those of Figure 1 of Hanchard et al. (2006). **Abbreviation:** Freq. — Frequency, Popu. — Population.

Selection Coefficient	Sample Size n	Recombination Rates ρ	Nominal Level α	Present Day Popu. Freq. of Derived Allele						
				0.1	0.2	0.4	0.6	0.8	0.9	
s=500	100	1.5#	0.05	0.9998	1.0000	1.0000	1.0000	1.0000	0.9998	
			0.01	0.9984	1.0000	1.0000	1.0000	1.0000	0.9998	
		3#	0.05	1.0000	1.0000	1.0000	1.0000	1.0000	0.9976	
			0.01	1.0000	1.0000	1.0000	1.0000	1.0000	0.9932	
		6#	0.05	0.9652	1.0000	1.0000	1.0000	1.0000	0.9602	
			0.01	0.8624	0.9998	1.0000	1.0000	1.0000	0.8576	
		9	0.05	0.9250	0.9994	1.0000	1.0000	1.0000	0.9994	0.8682
			0.01	0.7732	0.9986	1.0000	1.0000	1.0000	0.9944	0.6666
	60	1.5	0.05	0.9690	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
			0.01	0.9336	0.9996	1.0000	1.0000	1.0000	1.0000	0.9992
		3	0.05	0.9168	0.9990	1.0000	1.0000	1.0000	0.9998	0.9860
			0.01	0.8064	0.9970	1.0000	1.0000	1.0000	0.9990	0.9500
		6	0.05	0.7496	0.9898	1.0000	1.0000	1.0000	0.9844	0.7874
			0.01	0.5524	0.9522	1.0000	1.0000	1.0000	0.9420	0.5720
		9	0.05	0.6604	0.9806	1.0000	0.9998	1.0000	0.9702	0.5690
			0.01	0.4382	0.9282	0.9998	0.9998	1.0000	0.8864	0.3132